

8.1 Collecting Data – Overview

Population vs Sample

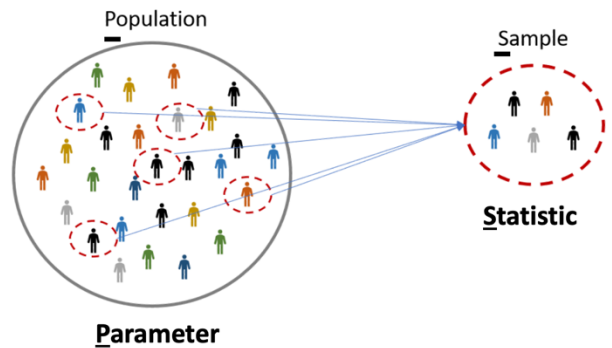
Ex) Lets say I want to figure out if Indiana is a cat or dog state.

Population – The particular group of interest in a study (the set of all individuals/objects of interest).

- Ex) Every person in ALL of Indiana.

Sample – A subset of individuals/objects from the population of interest.

- Ex) Everybody in Muncie ONLY.



Parameter vs Statistic

Population parameter – A fixed numerical value that describes the population.

- Ex) Overall percentage who prefer cats for IN
- Would have to take a **census** (ask everyone in the population) to know this value(or estimate it).

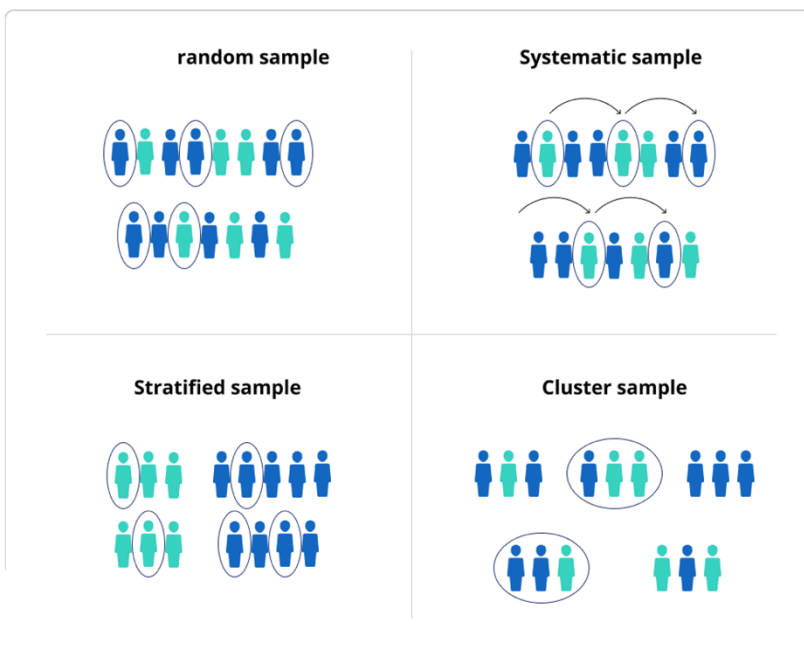
Sample Statistic – A numerical value that describes the sample that can vary from sample to sample.

- Ex) Percentage for Muncie (will be different than for Indy)

Sampling Techniques

GOAL: Representative Sample – A sample that has the same relevant characteristics as the population and does not favor one group of the population over another.

Matches / resembles the population



Convenience sample



1) Random Sample – A random sample is one in which every member of the population has an equal chance of being selected.

- This is generally desirable but can be difficult to achieve.

2) Stratified Random Sample – Dividing population into homogeneous (similar characteristics) groups.

This guarantees the sample is representative!

- 1) Stratify the population – Divide the population into similar groups (e.g. based on age or gender).
- 2) Take random sample from each group (strata).
- 3) Combine the groups from each strata to form your sample.

3) Cluster Sample – Dividing population into mini-populations. This gives us an unbiased sample and is often a more practical / affordable method.

- 1) Split the population into representative groups called clusters (resemble overall population).
- 2) Use random sampling to select several clusters.
- 3) Perform a census of each selected (collect data from every member).

4) Systematic Sample – Selecting every n^{th} member of the population.

5) Convenience Sample – Include individuals who are convenient to sample (for the researcher); AVOID!

- The group may not be representative of the population → Frequently ends in biased results.

Examples: Describe how you could obtain a sample to answer each question below using each of the following types of sampling methods listed below.

Example 1

Scenario: I wish to determine the proportion of the MATH 125 class that has a Mac laptop.

- Random Sample:

Randomly ask 10 students from class

- Stratified Random Sample:

Randomly sample 5 students from the list of Freshman and Sophomores respectively

- Cluster Sample:

Randomly sample 3 tables; ask everyone in that (take a census of the) table

- Systematic Sample:

Ask every 4th student of the roster

- Convenience Sample:

Ask the 5 people closest to me

Example 2

Scenario: You are tasked with conducting a survey to answer the question, “What is the favorite subject of students who attend East High School?”

- Random Sample:

Get a list of all student names and randomly select 10 names (could number each name and randomly generate 10 numbers)

- Stratified Random Sample:

Divide students by grade level OR by social group (band, football, etc.) and then randomly sample within each group

(Students are the same WITHIN each group, but different ACROSS groups)

- Cluster Sample:

Randomly sample 5 classrooms OR buses, and then ask every student in each.

(Should be a mix of students in each cluster, all clusters \approx same)

- Systematic Sample:

Ask every 5th student that arrives in the morning.

Get a list of all students names and select every 10th name.

- Convenience Sample: (AVOID -> Biased results)

Walk through hallway and ask first 20 students you pass.

(Bad because maybe I am by the chem lab and only ask chem students; their opinions might not match the overall student body's opinion)